SYSTEM AND METHOD FOR MALICIOUS CODE DETECTION

FIELD OF THE INVENTION

[01] The present invention relates to computer security and more particularly to a system and method for malicious code detection.

BACKGROUND OF THE INVENTION

[02] Malicious code is software that is designed to damage a computer system or its data or to prevent the computer system from being used in its normal manner. Also termed "malware," malicious code includes viruses, Trojan horses, worms, and malicious active content. A virus is a particularly pernicious kind of malicious code, capable of attaching itself to disks or other files and replicating itself repeatedly, typically without user knowledge or permission. Some viruses display symptoms, and some viruses damage files and computer systems, but neither symptoms nor damage is essential in the definition of a virus. A non-damaging virus is still a virus, yet even non-damaging viruses are considered malicious if they consume valuable computer resources without permission.

[03] Some viruses propagate by attaching themselves to files so that executing an infected file also causes the virus to execute. The virus then hooks into the operating system to infect other computer files as they are opened, modified or created. Before the popularity of the Internet, viruses were most commonly spread by sharing floppy disks that have been infected or that contain infected files. The recent, explosive growth of the Internet has increased the opportunities for spreading malicious code quickly throughout the world, for example, through infected files attached to electronic mail messages. When the email recipient executes an infected email attachment, the virus is propagated to yet another computer system.

[04] To combat viruses and other kinds of malicious code, vendors have begun to offer anti-virus software that scans incoming files and other content for embedded viruses, Trojan horses, malicious document macros, and worms. The incoming content that is scanned typically includes attachments to an email message, the body of the email message itself, and scripts downloaded via HTTP. Such anti-virus software typically employs a proprietary catalog of viral signatures, which are often simple string of bytes that are expected to be found in every instance of particular viruses. Usually, different viruses have different signatures, and anti-virus scanners use signatures to locate specific viruses.

[05] There are a large variety of viruses and other kinds of malicious code thriving on the Internet, but no single anti-virus scanner has 100% coverage of the known viruses. Each anti-virus scanner has its own set of viruses that the anti-virus scanner can detect, and many anti-virus scanners can detect viruses that are unknown to other anti-virus scanners on the market. Therefore, incomplete coverage of known viruses is a problem with individual anti-virus scanners.

[06] Accordingly, attempts have been made to improve virus coverage by employing a variety of different anti-virus scanners. One example is the VIRUS CONTROL CENTRE™, which is currently offered from MessageLabs™ and is described at the http://www.messagelabs.com web site. The VIRUS CONTROL CENTRE™ product comprises a cluster of control towers that are populated with a plurality of scanning mail servers, a switch, and a load distributor. All incoming email is redirected to a control tower for initial processing and scanning. After being delivered to a control tower, the email is directed to a particular scanning mail server, which executes three different types of commercial anti-virus scanners on the email. If the email is "clean," then the email is permitted to continue to its ultimate destination. Otherwise, the email is quarantined for 30 days and then destroyed.

[07] This approach, however, suffers from several disadvantages, particularly in terms of latency. Latency is the delay imposed by scanning for viruses. For example, if each anti-virus scanner on the scanning mail server takes 400 ms to process an average email, then the latency imposed by the three anti-virus scanners is 1.2 seconds.

[08] Although a 1.2 second latency may appear to be small at first blush, it is unacceptably large for interactive traffic such as surfing the World Wide Web. Email is not the only vector for transmitting malicious code, viruses can also be downloaded in web pages sent by the hypertext transfer protocol (HTTP) or in files sent by the file transfer protocol (FTP). If a user had to wait 1.2 seconds every time to see a new web page, the user would quickly become frustrated and seek less secure ways of accessing the Internet. On the other hand, a latency of about 0.5 seconds is still acceptable to most users.

[09] Therefore, there is a need for a malicious code detection system and methodology with the good anti-viral coverage of multiple anti-virus scanners but characterized by the low latency commensurate with that of a single anti-virus scanner.

SUMMARY OF THE INVENTION

[10] These and other needs are addressed by the present invention, in which incoming content scanned in parallel by different anti-virus software on separate processors in a multi-processor or multi-computer configuration. By scanning incoming content in parallel on separate processors, the latency of scanning the content is reduced to that of only one of the anti-virus scanners plus a small amount of overhead. For example, if three anti-virus scanners operating in parallel have an average latency of 400 ms each, the overall latency due to the parallel operation is not 1.2 ms but 400 ms plus a 10% overhead for a 440 ms overhead, which is acceptable to users surfing the World Wide Web.

[11] Accordingly, one aspect of the invention relates to a system and methodology for malicious code detection. The system includes a front-end processor, multiple scanning computer systems, and a detection management system. The multiple scanning computer systems are configured for scanning content for malicious code and generating an alarm when the content contains malicious code. The front-end processor, which is coupled to the scanning computer systems, receives a flow of content (including, for example, email message bodies, email attachments, HTTP or FTP files) from an external network, such as the Internet, and distributes copies of the flow to each of the scanning computer systems in parallel for scanning. The detection management system, also coupled to the scanning computer systems, employs a countermeasure on the flow if at least one of the scanning computer systems generates the alarm.

[12] Another aspect of the invention relates to a malicious code detection system and methodology that includes a remote site detection system configured for detecting malicious code in incoming network traffic based on signatures of malicious code. In this system, multiple scanning computer systems configured to execute anti-virus scanning software having different coverage of malicious code for scanning content for malicious code and generating an alarm when the content contains malicious code. A front-end processor, coupled to the scanning computer systems, is configured for receiving a flow of content (including, e.g., email attachments, an email message body, a hypertext markup file or a transferred file) from an external network and distributing copies of the flow to each of the scanning computer systems in parallel for scanning. A detection management system, coupled to the scanning computer systems, is configured for creating a signature of a piece of malicious code detected by at least one of the scanning computer systems detected in the flow when at least one of the scanning computer generates an alarm on the piece of malicious code and causing the signatures stored at the remote site detection system to be updated to include the signature of the detected piece of malicious code. The detection management system also employs a

countermeasure on the flow, which includes blocking the flow, quarantining the flow, or informing the recipient of the flow of the malicious code.

[13] Still another aspect of the present invention pertains to a front-end system, its method, and its software. The front-end system is coupled to an external network and multiple scanning computer systems and is configured for receiving a flow of content (including a hypertext markup file or a transferred file) from the external network, duplicating the flow to produce multiple copies of the flow, and distributing a copy to each of the multiple scanning computer systems in parallel.

[14] Yet another aspect of the present invention involves a malicious code detection cluster and its methodology that includes an internal network coupled to a front-end processor, a detection management system, and multiple scanning computer systems. The multiple scanning computer systems are configured for receiving copies of a flow of content (including, for example, email messages, their attachments and bodies, a hypertext markup file or a transferred file), executing anti-virus scanning software with different coverages to scan the copies of the flow in parallel for malicious code, and transmitting an alarm to the detection management system when the flow contains malicious code as detected by at least one of the multiple scanning computer systems.

[15] An additional aspect of the present invention relates to a detection management system, method, and software that are used in conjunction with multiple scanning computer systems. An alarm is received from one of the multiple scanning computer systems when a flow of content (including a hypertext markup file or a transferred file) scanned by the scanning computer systems in parallel contains malicious code. A countermeasure is employed on the flow if at least one of the scanning computer systems generates an alarm on a piece of malicious code. In one embodiment, a signature of a piece of malicious code detected by at least one of the scanning computer systems in the flow is created when at least one of the scanning computer generates an alarm on the

piece of malicious code. Signatures stored at a remote site detection system are then updated to include the created signature.

[16] Still other objects and advantages of the present invention will become readily apparent from the following detailed description, simply by way of illustration of the best mode contemplated of carrying out the invention. As will be realized, the invention is capable of other and different embodiments, and its several details are capable of modifications in various obvious respects, all without departing from the invention. Accordingly, the drawing and description are to be regarded as illustrative in nature, and not as restrictive.

BRIEF DESCRIPTION OF THE DRAWINGS

[17] The present invention is illustrated by way of example, and not by way of limitation, in the figures of the accompanying drawings and in which like reference numerals refer to similar elements and in which:

[18] FIG. 1 is a diagram of one embodiment of a malicious code detection system in accordance with the present invention.

[19] FIG. 2 is a flow chart illustrating the operation of one embodiment of a malicious code detection methodology in accordance with the present invention.

[20] FIG. 3 depicts a computer system that can be used to implement various aspects of an embodiment of the present invention.

DESCRIPTION OF THE PREFERRED EMBODIMENT

[21] A system, method, and software for malicious code detection are described. In the following description, for the purposes of explanation, numerous specific details are set forth in order to provide a thorough understanding of the present invention. It will be apparent, however, to one skilled in the art that the present invention may be practiced without these specific details. In other instances, well-known structures and devices are

shown in block diagram form in order to avoid unnecessarily obscuring the present invention.

MALICIOUS CODE DETECTION SYSTEM

[22] FIG. 1 depicts one embodiment of a malicious code detection system in accordance the present invention, whose operation is described in conjunction with FIG. 2. The malicious code detection system, which is suitable for deployment by Internet Service Providers (ISPs), network service providers, and corporate information systems departments for responsible for the employees' email and web browsing, is coupled to the Internet or other external network 100 for receiving content such as web pages 102, email messages 104, attachments 106 to the email messages 104, and raw files 108 transmitted over a file transfer protocol such as FTP. Since this content is received from an external network, they may contain viruses or other kinds of malicious code and therefore need to be scanned by the malicious code detection system.

[23] Accordingly, a front-end processor 110 is coupled to the external network 100 as part of, or in conjunction with other external interface equipment (not shown), such as firewalls and load balancers. The front-end processor 110 is a computer system that is configured for receiving a "flow" of one or more of the incoming content 102, 104, 106, and 108 from the external network 110 (FIG. 2, step 200). When each flow of content is received, it is assembled into the appropriate end result (FIG. 2, step 202). For example, files received over Simple Mail Transfer Protocol (SMTP) are assembled into an email body and series of email attachment. Files received over FTP are assembled into a series of raw files, whether binary or ASCII. The HTTP traffic is assembled into a series of web pages.

[24] The front-end processor 110 is responsible for duplicating the assembled flows for distribution to multiple scanning computer systems 122, 124, and 126 in parallel (FIG. 2, step 204). Although three multiple scanning computer systems 122, 124, and

126 are illustrated in FIG. 1, the present invention is not so limited and any number may be used. In addition, in certain implementations such as those using the CISCO CATALYST™ family switch, the front-end processor 110 may perform other functions such as load balancing. In one embodiment, the front-end processor 110 and the multiple scanning computer systems 122, 124, and 126 are coupled to a common, high-speed internal network 120, such as a fast Ethernet™ network, but in other embodiments dedicated connections may be employed between the front-end processor 110 and each of the multiple scanning computer systems 122, 124, and 126 instead.

[25] Each of the multiple scanning computer systems 122, 124, and 126 includes a cluster of one or more processors running anti-virus software for scanning a corresponding copy of the flow for viruses and other kinds of malicious code (FIG. 2, stage 206). Different anti-virus software, obtained from different software vendors and having different coverage of known viruses, are employed to obtain an anti-viral coverage that is better than any single anti-virus software product. In high-performance implementations, extra scanning computer systems are deployed to process the flow at a higher throughput. In these implementations, the front-end processor 110 preferably performs load balancing to ensure that each of the scanning computer systems is fully utilized.

[26] When any of the multiple scanning computer systems 122, 124, and 126 detects a virus or other kind of malicious code in the flow, the detecting scanning computer system generates an alarm, which is sent to a detection management process executing on a detection management system 130 (FIG. 2, stage 208). Preferably, the detection management system 130 is also coupled to the internal network 120 and is deployed on a separate computer system. However, other implementations are possible; for example, the detection management system 130 may be put on the same computer system as the front-end processor 110 or one of the multiple scanning computer systems 122, 124, and 126.

[27] The detection management system 120 integrates any possible alarms received from the multiple scanning computer systems 122, 124, and 126 (FIG. 2, step 210) and checks whether an alarm was generated for a particular flow (FIG. 2, step 212). If the detection management system 130 does not receive an alarm from any of the multiple scanning computer systems 122, 124, and 126 for a particular flow, then the flow is directed to its ultimate destination, for example, to a user computer connected to an ISP or to a corporate intranet (FIG. 2, step 214).

[28] On the other hand, if the detection management system 130 does receive an alarm from any of the multiple scanning computer systems 122, 124, and 126 for a particular flow, then an appropriate countermeasure is executed on the flow (FIG. 2, step 216). Various countermeasures may be employed and include any one or more of the following: destroying the flow, quarantining the flow in a safe directory for a period of time such as thirty days for possible study, and emailing a separate message or embedding a message to the recipient and/or sender informing the person that the email message contained a virus. For an HTTP web page, an appropriate message may be embedded into the web page, e.g. by appropriate HTML or other markup or by Javascript™ or other scripting language instructions, informing the surfer of the virus in the web page.

[29] Because multiple scanning computer systems 122, 124, and 126 inspect the flows for malicious code in parallel, the latency of the system is limited to the latency of the slowest one of the multiple scanning computer systems 122, 124, and 126 plus some overhead for duplicating and distributing the flow and integrating the alarms. Consequently, the total latency is commensurate with that of one anti-virus scanner, not the sum total of all the different anti-virus scanners as in prior approaches, while the anti-viral coverage is a superset of the anti-virus scanners. This solution is particularly advantageous for people using a browser to access the World Wide Web to view web pages or transfer files that may contain viruses. The latency is at an acceptable half

second, instead of the unacceptably slow 1.2 seconds of the serial anti-virus scan approach.

DYNAMIC ALLOCATIONS TO REMOTE SITES

[30] Some customers may wish to take advantage of the improved coverage of multiple anti-virus scanners but cannot afford the hardware and software costs associated with the full solution. Accordingly, one aspect of the present invention involves a mechanism for integrating the broad coverage obtained from the multiple anti-virus scanner solution for use by remote sites that can only afford one anti-virus scanner.

[31] In an embodiment of this aspect, as illustrated in FIG. 1, the detection manager system 130 is also coupled to a relational or other kind of database 132. The database 132 stores rules for creating signatures of detected viruses. Thus, as viruses are detected by the detection manager system 130 in response to alarms generated by the multiple scanning computer systems 122, 124, and 126, the detection manager system 130 creates a signature that identifies the virus or other kind of malicious code, such as a Trojan horse, worm, etc. (FIG. 2, step 218).

[32] Periodically, or upon detection of a new virus, the detection manager system 130 transmits the new signatures to the remote site scanning system 140 to augment the catalog of signatures stored at the remote site scanning system 140 (FIG. 2, step 220). As a result, the remote site scanning system 140 is updated to include the signatures of the latest viruses.

[33] Coverage of live viruses that were detected by the detection manager system 130 is particularly beneficial and cost effective for a small remote site scanning system 140. Although the native anti-virus scanner at a small remote site scanning system 140 is not as broad at the aggregate coverage of the multiple scanning computer systems 122, 124, and 126, the volume of the traffic through the small remote site scanning system 140 is typically much smaller than the volume of the traffic through the high-performance front-

end processor 110. Consequently, a virus during an outbreak is more likely to be transmitted through the higher-volume front-end processor 110 before reaching the lower-volume small remote site scanning system 140. In this common situation, the system comprising the front-end processor 110, the multiple scanning computer systems 122, 124, and 126, and the detection manager system 130 are able to preemptively identify a virus and add its signature to the small remote site scanning system 140, well before the virus is actually transmitted to the small remote site scanning system 140. Thus, the small remote site scanning system 140 is able to take advantage of the broader anti-virus scanning coverage of the multiple scanning computer systems 122, 124, and 126 without the comparable investment in hardware resources.

HARDWARE OVERVIEW

[34] FIG. 3 is a block diagram that illustrates a computer system 300 upon which an embodiment of the invention may be implemented. Computer system 300 includes a bus 302 or other communication mechanism for communicating information, and a processor 304 coupled with bus 302 for processing information. Computer system 300 also includes a main memory 306, such as a random access memory (RAM) or other dynamic storage device, coupled to bus 302 for storing information and instructions to be executed by processor 304. Main memory 306 also may be used for storing temporary variables or other intermediate information during execution of instructions to be executed by processor 304. Computer system 300 further includes a read only memory (ROM) 308 or other static storage device coupled to bus 302 for storing static information and instructions for processor 304. A storage device 310, such as a magnetic disk or optical disk, is provided and coupled to bus 302 for storing information and instructions.

[35] Computer system 300 may be coupled via bus 302 to a display 312, such as a cathode ray tube (CRT), for displaying information to a computer user. An input device 314, including alphanumeric and other keys, is coupled to bus 302 for communicating

information and command selections to processor 304. Another type of user input device is cursor control 316, such as a mouse, a trackball, or cursor direction keys for communicating direction information and command selections to processor 304 and for controlling cursor movement on display 312. This input device typically has two degrees of freedom in two axes, a first axis (e.g., *x*) and a second axis (e.g., *y*), that allows the device to specify positions in a plane.

[36] The invention is related to the use of computer system 300 for aspects of malicious code detection. According to one embodiment of the invention, various aspects of malicious code detection are provided by computer system 300 in response to processor 304 executing one or more sequences of one or more instructions contained in main memory 306. Such instructions may be read into main memory 306 from another computer-readable medium, such as storage device 310. Execution of the sequences of instructions contained in main memory 306 causes processor 304 to perform the process steps described herein. One or more processors in a multi-processing arrangement may also be employed to execute the sequences of instructions contained in main memory 306. In alternative embodiments, hard-wired circuitry may be used in place of or in combination with software instructions to implement the invention. Thus, embodiments of the invention are not limited to any specific combination of hardware circuitry and software.

[37] The term "computer-readable medium" as used herein refers to any medium that participates in providing instructions to processor 304 for execution. Such a medium may take many forms, including but not limited to, non-volatile media, volatile media, and transmission media. Non-volatile media include, for example, optical or magnetic disks, such as storage device 310. Volatile media include dynamic memory, such as main memory 306. Transmission media include coaxial cables, copper wire and fiber optics, including the wires that comprise bus 302. Transmission media can also take the form of acoustic or light waves, such as those generated during radio frequency (RF) and

infrared (IR) data communications. Common forms of computer-readable media include, for example, a floppy disk, a flexible disk, hard disk, magnetic tape, any other magnetic medium, a CD-ROM, DVD, any other optical medium, punch cards, paper tape, any other physical medium with patterns of holes, a RAM, a PROM, an EPROM, a FLASH-EPROM, any other memory chip or cartridge, a carrier wave as described hereinafter, or any other medium from which a computer can read.

[38] Various forms of computer readable media may be involved in carrying one or more sequences of one or more instructions to processor 304 for execution. For example, the instructions may initially be borne on a magnetic disk of a remote computer. The remote computer can load the instructions into its dynamic memory and send the instructions over a telephone line using a modem. A modem local to computer system 300 can receive the data on the telephone line and use an infrared transmitter to convert the data to an infrared signal. An infrared detector coupled to bus 302 can receive the data carried in the infrared signal and place the data on bus 302. Bus 302 carries the data to main memory 306, from which processor 304 retrieves and executes the instructions. The instructions received by main memory 306 may optionally be stored on storage device 310 either before or after execution by processor 304.

[39] Computer system 300 also includes a communication interface 318 coupled to bus 302. Communication interface 318 provides a two-way data communication coupling to a network link 320 that is connected to a local network 322. For example, communication interface 318 may be an integrated services digital network (ISDN) card or a modem to provide a data communication connection to a corresponding type of telephone line. As another example, communication interface 318 may be a local area network (LAN) card to provide a data communication connection to a compatible LAN. Wireless links may also be implemented. In any such implementation, communication interface 318 sends and receives electrical, electromagnetic or optical signals that carry digital data streams representing various types of information.

[40] Network link 320 typically provides data communication through one or more networks to other data devices. For example, network link 320 may provide a connection through local network 322 to a host computer 324 or to data equipment operated by an Internet Service Provider (ISP) 326. ISP 326 in turn provides data communication services through the worldwide packet data communication network, now commonly referred to as the "Internet" 328. Local network 322 and Internet 328 both use electrical, electromagnetic or optical signals that carry digital data streams. The signals through the various networks and the signals on network link 320 and through communication interface 318, which carry the digital data to and from computer system 300, are exemplary forms of carrier waves transporting the information.

[41] Computer system 300 can send messages and receive data, including program code, through the network(s), network link 320, and communication interface 318. In the Internet example, a server 330 might transmit a requested code for an application program through Internet 328, ISP 326, local network 322 and communication interface 318. In accordance with the invention, one such downloaded application provides for malicious code detection as described herein. The code may be executed by processor 304 as it is received, and/or stored in storage device 310, or other non-volatile storage for later execution. In this manner, computer system 300 may obtain application code in the form of a carrier wave.

[42] Accordingly, a system, methodology, and software for detection of malicious code is described, in which content from an external network is scanned by software multiple scanning computer systems in parallel. Latency is reduced from the sum of the delays introduced by all the malicious code scanners to be commensurate with the delay of one of the malicious code scanner, without comprising coverage. Furthermore, the benefits of the increased coverage can be transmitted to remote site scanning systems, without the need for additional hardware costs.

[43] While this invention has been described in connection with what is presently considered to be the most practical and preferred embodiment, it is to be understood that the invention is not limited to the disclosed embodiment, but on the contrary, is intended to cover various modifications and equivalent arrangements included within the spirit and scope of the appended claims.